IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 January 2024, accepted 1 February 2024, date of publication 26 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370439

RESEARCH ARTICLE

Stochastic Modeling of the Set-Membership-Sign-NLMS Algorithm

JOSÉ V. G. DE SOUZA[1], FELIPE DA R. HENRIQUES[2], NEWTON N. SIQUEIRA[3], LUÍS TARRATACA[2], FABIO A. A. ANDRADE[4], (Senior Member, IEEE), AND DIEGO B. HADDAD[2], (Member, IEEE)

[1]Federal Center for Technological Education, Rio de Janeiro 20271-110, Brazil
[2]Federal Center for Technological Education, Petrópolis 20271-110, Brazil
[3]Federal Center for Technological Education, Nova Iguaçu 20271-110, Brazil
[4]Department of Microsystems, University of South-Eastern Norway, Horten, 3199 Borre, Norway

Corresponding author: Fabio A. A. Andrade (fabio@ieee.org)

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) under Grant 001, and in part by Fundação Carlos Chagas Filho de Amparo ã Pesquisa do Estado do Rio de Janeiro (FAPERJ).

ABSTRACT This paper proposes the Set-Membership sign-NLMS (SM-sign-NLMS) adaptive filter, which combines the ability for data censoring (offered by Set-Membership schemes) with robustness against impulsive noise (provided by signed schemes). The algorithm can present a much lower steady-state probability of update than the standard SM-NLMS algorithm when impulsive noise is present in the system. It is derived from a local deterministic optimization problem modulated by a minimum disturbance cost function combined with a bounded error criterion. Several stochastic models are proposed in order to extract insights and a time-variant step size extension of the algorithm. The first of them, based on energy conservation arguments, leads to a fixed-point analytic equation whose solution predicts the asymptotic performance of the algorithm. Further, a transient analysis based on a statistical decoupling of the radial and (discrete) angular distributions of the input vector is derived. Based on such an analysis, an efficient time-variant step-size version of the algorithm is proposed. Additionally, such an analysis is also utilized to obtain a fixed-point formula whose solution describes the asymptotic performance when the unknown plant that the filter intends to match varies according to a first-order Markovian model. Lastly, a novel stochastic model is advanced for the description of the algorithm learning behavior under a deficient-length scenario for a white input signal, which provides some insights about the asymptotic performance of the algorithm. The findings are confirmed by extensive simulations.

INDEX TERMS Set-membership, adaptive filtering, data censoring, computational complexity, stochastic models.

I. INTRODUCTION

Signals are often contaminated by unwanted artifacts and noise that affect the performance of adaptive filters. Such filters, originally derived from the optimal prediction and filtering method for solving the Wiener-Hoff equation, play a relevant role in advanced signal processing and control schemes [1], [2]. The optimization problem they solve is intrinsically related to their learning features, such as stability, computational cost and robustness against impulsive noise that may be found in the measurement noise. The latter phenomenon may occur due to a plethora of possible causes, ranging from a large number of drum shrimp family organisms in hydroacoustic channels [3] to double-talk in acoustic echo cancellation systems [4] and atmospheric phenomena in telecommunication systems [5]. Applications that have to deal with impulsive interferences range from echo cancellation to signal prediction and location tracking [6].

In this paper, a novel Set-membership signed-error normalized LMS (SM-sign-NLMS) algorithm is derived

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. Garcí-a-Ferández.

from a minimum disturbance optimization problem. The algorithm operates, in some scenarios, in an attractive point of the ubiquitous trade-off between convergence rate, asymptotic performance and computational burden. Namely, the method blends the reduction of computational cost from the Set-membership approach (due to its data censoring capabilities) with the robustness against impulsive noise provided by the signed-error strategies.[1] It is noteworthy that such a reduction is crucial in applications that demand thousands of adaptive taps (*e.g.*, in the acoustic echo cancellation task [7]). Furthermore, the data-dependent selective update of the Set-membership schemes evaluates the incoming data in terms of their contribution to the estimation procedure [8].

One weakness of traditional adaptive filtering schemes with ℓ_2-norm-based cost functions is their vulnerability to disturbances such as the ones that occur with impulsive noise [9], [10]. In this context, signed variants receive great attention, since the occurrence of impulsive noise causes high fluctuations of the error signal, resulting in updates in the wrong direction (or even divergence) in traditional adaptive schemes [9], [11]. The signed-error variants utilize nonlinear correlation multipliers [12] that can offer robustness against impulsive noise. Typically, such a disturbance does not induce a high misalignment in the adaptive estimator when the sign of the error is used, instead of the error signal [13].

This paper advances a framework that is able to furnish an algorithm that combines both Set-membership and signed approaches. The proposed algorithm eliminates the need for parameter selection by leveraging prior knowledge of the noise, a feature shared by both correntropy-based and the least mean p-th power algorithms [14], [15]. Despite its simple update equation and low computational burden, the learning behavior of the advanced algorithm is very sophisticated (a feature it shares with adaptive algorithms in general). This fact demands the right level of theorizing craft, in order to comply with the demands of both accurate predictions and extraction of relevant insights about different aspects of the algorithm learning mechanism.

Models that allow for more hypothesis usually translate into simpler equations that lose adherence to the original data. Accordingly, the model needs to be precisely calibrated, such that concise equations allowing for insight extraction are obtained but, at the same time, they are not too simplified so as to provoke a lack of adherence to real behaviour. This leads to the choice of a specific model that operates in a set point of the explanation *versus* prediction dimensions. Since each model has its weaknesses, distinct stochastic models are employed in this paper, in order to attain a broader spectrum of insights regarding the performance of the proposed algorithm. Those theoretically-based perceptions have attracted attention from the scientific community, since they offer *i)* useful guidelines for the algorithm designer; *ii)* novel relationships with remaining adaptive schemes; *iii)* interpretable cause-effects relations; *iv)* new ways of explaining some phenomenon in a more transparent manner; and *v)* suggest further questions or generalizations of practical interest.

This paper is structured as follows. The novel algorithm, that combines the advantageous features of Set-membership and signed-variants schemes, is derived in Section II through a deterministic and purely local optimization problem. A fixed-point equation whose solution estimates its asymptotic mean square performance is obtained through energy-conservation arguments in Section III. A transient analysis that simplifies the joint statistics of the input vector is advanced in Section IV to obtain a modal description of its transient behavior. Since the model for the input excludes its Gaussianity (except possibly for the radial distribution), which does not allow the utilization of the Price theorem, the recently proposed *Price heuristics* [16] is adopted in the model. Such a model is utilized in Section V for proposing a practical variable step-size (VSS) scheme tailored for the algorithm, which does not requires additional adjustable parameters (*i.e.*, a novel *nonparametric* VSS is proposed, see [17]). The advanced VSS scheme significantly improves the rate convergence compared to its original fixed step-size version. The analysis of Section IV is extended in Section VI to describe its steady-state performance when the unknown plant varies according to a first-order Markovian model. In Section VII-B, the asymptotic mean squared deviation in the case of a deficient-length adaptive filter is also obtained as a solution of two coupled fixed-point equations, utilizing a stochastic model based on a recursion of the autocorrelation matrix of the deviation coefficients.

A. MATHEMATICAL NOTATION

Throughout this paper, vector and matrices are represented with lowercase and uppercase bold fonts, respectively, while scalars are denoted by italics. $\|\mathbf{x}\|^2$ denotes the squared Euclidean norm of vector $\mathbf{x}$, and $\mathbb{E}[\cdot]$ is the expectation statistical operator. $(\cdot)^T$ denotes transpose and Tr[$\mathbf{A}$] is the trace of matrix $\mathbf{A}$. All vectors are of column type. The function sign(x) determines the output value based on whether the input argument x is positive, negative, or zero, returning $+1$ for positive, -1 for negative, and 0 for zero input. Symbol $\sim$ indicates that a random variable (or random vector) is distributed according to the same probability density function as another given random variable. Prob$\{\mathcal{A}\}$ represents the probability of event $\mathcal{A}$ occurring. Expression $\nabla_{\mathbf{w}} f$ computes the gradient of the *scalar* function $f(\mathbf{w})$ w.r.t. the vector $\mathbf{w}$.

II. THE SM-SIGN-NLMS ALGORITHM

In a system identification setting, Set-membership Filtering (SMF) approaches exploit the hypothesis of a bounded noise process $\nu(k)$ immersed in a linear-in-the-parameters model [18]:

$$d(k) = \left[\mathbf{w}^{\star}\right]^T \mathbf{x}(k) + \nu(k), \qquad (1)$$

[1] Signed-error variants may reduce the computational burden of the filter, but this is not the case with our method, due to its normalization procedure.

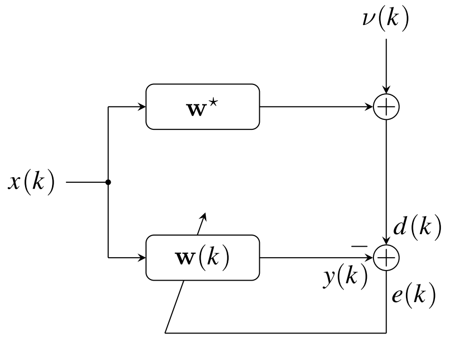


FIGURE 1. Block diagram of the structure of an adaptive filtering algorithm applied to a system identification task.

where $d(k) \in \mathbb{R}$ is the reference signal, $\mathbf{w}^\star \in \mathbb{R}^N$ contains the ideal (and unknown) set of coefficients the adaptive filter intends to estimate and $\mathbf{x}(k)$ comprises N consecutive samples of the input signal $x(k)$ at the k-th iteration:

$$\mathbf{x}(k) \triangleq \left[x(k)\; x(k-1)\; \ldots\; x(k-N+1)\right]^T, \tag{2}$$

where a transversal structure is assumed.

In each iteration, the adaptive filter generates an output sample $y(k) \in \mathbb{R}$, computed through an inner product:

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k), \tag{3}$$

where $\mathbf{w}(k) \in \mathbb{R}^N$ contains the adaptive taps at the k-iteration:

$$\mathbf{w}(k) \triangleq \left[w_0(k)\; w_1(k)\; \ldots\; w_{N-1}(k)\right]^T, \tag{4}$$

which should be adapted in an iterative manner by a specific adaptation rule. The difference between the reference signal and the filter output is the error signal $e(k) \in \mathbb{R}$:

$$e(k) \triangleq d(k) - \mathbf{w}^T(k)\mathbf{x}(k). \tag{5}$$

The overall system identification task this paper focuses on is depicted in Figure 1.

The SMF aims to guarantee a prescribed bound on the magnitude of the error within the relevant time frame. In this sense, any update of the adaptive vector that provides an absolute value of the error less than the adjustable bound $\overline{\gamma}$ is considered to be a feasible solution for the iterative learning procedure. Moreover, such a methodology bounds the worst-case error achieved by the filter, with prior error-bound specification [8], [19]. Typically, the threshold $\overline{\gamma}$ depends on the variance of the additive noise σ_ν^2 through [20]

$$\overline{\gamma} = \sqrt{\tau\sigma_\nu^2}, \tag{6}$$

where $\tau \in \mathbb{R}_+$ is an adjustable parameter. The effectiveness of SMF methodologies is tied to the specification of $\overline{\gamma}$, a task that may be challenging in real-world scenarios owing to the lack of knowledge regarding the environment and its dynamic intricacies [21]. In spite of that, it should be noted that the literature in general shows that SM algorithm outperform their non-SM counterparts [22].

Consider $\mathcal{S}$ as the dataset containing the pairs $\{\mathbf{x}(k), d(k)\}$ available for the learning scheme. The constraint set $\mathcal{H}_n$ is the set of adaptive weight vectors that are consistent with

$$\mathcal{H}_k \triangleq \{\mathbf{w} \in \mathbb{R}^N : |d(k) - \mathbf{w}^T(k)\mathbf{x}(k)| \leq \overline{\gamma}\}, \tag{7}$$

which defines a region enclosed by parallel hyperplanes and suggests the usage of more constraint-sets in the update mechanism [18], [23]. The standard SM-NLMS algorithm solves the following optimization problem [24]:

$$\begin{aligned} &\min_{\mathbf{w}(k+1)} \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2 \\ &\text{s.t. } \mathbf{w}(k+1) \in \mathcal{H}_k, \end{aligned} \tag{8}$$

where the output error is bounded by a prespecified bound (see (7)). The non-relaxed solution of (8) is

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \dfrac{\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2}\left[e(k) - \overline{\gamma}\right], & \text{if } |e(k)| > \overline{\gamma} \\ \mathbf{w}(k), \quad \text{otherwise}, \end{cases} \tag{9}$$

which demands fewer updates to reach steady state [24]. By utilizing the SMF, it becomes possible to lessen computational complexity in adaptive filtering, given that updates to filter coefficients happen only when the estimated error goes beyond the pre-established upper threshold [25].

The term $\|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2$ minimized in (8) derives from the conservative minimum disturbance principle, which assumes that the previous updates contain more information than the one brought by the current input vector, and therefore the current solution $\mathbf{w}(k)$ should be slightly perturbed in a reasonable learning process.

The standard sign-LMS algorithm is typically derived from a stochastic gradient procedure based on a ℓ_1-norm optimization [26]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \beta\nabla_{\mathbf{w}(k)}|e(k)| \tag{10}$$

$$\Rightarrow \mathbf{w}(k+1) = \mathbf{w}(k) + \beta\mathbf{x}(k)\text{sign}[e(k)], \tag{11}$$

where the update term $\beta\mathbf{x}(k)\text{sign}[e(k)]$ is a clipped function of the noise and the adaptive taps that replaces the error signal by its polarity [27], [28]. The hard limiter in Equation (11) complicates the stochastic modelling of the resulting algorithm, often requiring significant analytical ingenuity for a theoretical analysis to be conducted satisfactorily [29].

It is not trivial to design a linear-in-parameter filter whose space of feasible solutions for the updates combines the SMF data censoring capabilities and the robustness against impulsive noise offered by signed schemes. This is due to the fact that whereas the SMF approach employs the minimum disturbance criterion, the signed strategy usually adopts the stochastic gradient optimization method in its derivation. By taking the structure and constraints of the problem into account and using the conceptual links between the minimum disturbance principle and the stochastic gradient offered by

[30], [31], [32], and [33], this issue is circumvented in this paper by the following advanced optimization:

$$\min_{\mathbf{w}(k+1)} \frac{1}{2}\|\mathbf{w}(k+1)-\mathbf{w}(k)\|^2$$
$$\text{s.t. } e_{\mathrm{p}}(k) = \begin{cases} e(k), & \text{if } |e(k)| \le \overline{\gamma} \\ \left[1-\dfrac{\beta}{|e(k)|}\right]e(k), & \text{if } |e(k)| > \overline{\gamma}, \end{cases} \tag{12}$$

where $e_{\mathrm{p}}(k)$ denotes the *a posteriori* error, obtained *after* the update with the *current* data $\{d(k), \mathbf{x}(k)\}$:

$$e_{\mathrm{p}}(k) \triangleq d(k) - \mathbf{w}^T(k+1)\mathbf{x}(k). \tag{13}$$

Observe that the second constraint of (12) (*i.e.*, $e_{\mathrm{p}}(k) = \left[1-\frac{\beta}{|e(k)|}\right]e(k)$) imposes *diminishing returns* on the impact that the error $e(k)$ has on the update intensity. Such a feature is responsible for the desirable robustness against impulsive noise presented by the algorithm.

Theorem. The solution of (12) defines the following novel adaptive update:

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k), & \text{for } |e(k)| \le \overline{\gamma} \\ \mathbf{w}(k) + \beta\dfrac{\text{sign}[e(k)]}{\|\mathbf{x}(k)\|^2}\mathbf{x}(k), & \text{for } |e(k)| > \overline{\gamma}, \end{cases} \tag{14}$$

which combines the advantages of both Set-Membership and signed strategies.

Proof: Solving (12) by the Lagrange multipliers technique, the solution of (12) can be converted into the following unconstrained and equivalent optimization problem:

$$\mathcal{F}[\mathbf{w}(k+1)] = \frac{1}{2}\|\mathbf{w}(k+1)-\mathbf{w}(k)\|^2 + \lambda\left\{e_{\mathrm{p}}(k) - \left[1-\frac{\beta}{|e(k)|}\right]e(k)\right\}. \tag{15}$$

Zeroing the gradient of (15) leads to:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \lambda\mathbf{x}(k). \tag{16}$$

If $|e(k)| \le \overline{\gamma}$, the solution of (12) is trivial, since setting

$$\mathbf{w}(k+1) = \mathbf{w}(k) \tag{17}$$

complies with the constraint $e_{\mathrm{p}}(k) = e(k)$ and minimizes the term $\frac{1}{2}\|\mathbf{w}(k+1)-\mathbf{w}(k)\|^2$. This possibility imposes the condition $\lambda = 0$ in (16) and explains the advantageous reduction of the computational burden, since the update of the adaptive weight is avoided when the current data is redundant. Further, this implicit evaluation of the information content of the input data is also responsible for the data censoring capability of the resulting algorithm.

In the case $|e(k)| > \overline{\gamma}$, one may apply (16) in the second constraint of (12), leading to:

$$\lambda = \beta\text{sign}[e(k)]. \tag{18}$$

Bringing together equations (16), (17), and (18) takes one to (14). □

For our purposes of analytical description of the learning behavior of our novel method, a more adequate description of it is

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \beta\frac{\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2}f[e(k)], \tag{19}$$

where

$$f[e(k)] = \begin{cases} 0, & \text{for } |e(k)| \le \overline{\gamma}, \\ \text{sign}[e(k)] & \text{for } |e(k)| > \overline{\gamma}. \end{cases} \tag{20}$$

The proposed algorithm, as well as typical Set-membership algorithms, utilizes a metric projection onto a closed convex set (convex projection). Unfortunately, the nonlinearity of the convex projection poses a challenge to theoretical analysis. In the following sections, distinct stochastic models are employed in order to extract insights about the algorithm performance.

From a computational complexity perspective, the number of scalar sums/subtractions (resp. scalar multiplications) of the advanced algorithm is $2N+2$ (resp. $2N+1$), when there is an update. Only one scalar division for update is demanded. Note that such a complexity is very similar to the LMS algorithm, which requires $2N$ scalar sums/subtractions and $2N+1$ scalar multiplications per iteration. In fact, in practice the computational burden of the advanced algorithm is even less than this, since it does not update the weight vector in all iterations.

III. STEADY-STATE ANALYSIS

The SM-sign-NLMS algorithm involves two nonlinearities: the data normalization [34] and the error nonlinearity [35]. Treating these nonlinearities at the same time makes the performance analysis difficult [36]. Thus, some stochastic assumptions will be introduced to overcome such hurdle.

In this section, energy-conservation arguments, that comprise one of the most powerful analyses (based on energy conservation arguments), are adopted for the purpose of obtaining a prediction for the steady-state performance of the algorithm. Consider the deviation vector described by

$$\tilde{\mathbf{w}}(k+1) \triangleq \mathbf{w}^\star - \mathbf{w}(k). \tag{21}$$

where $\mathbf{w}^\star \in \mathbb{R}^N$ is the optimal solution.

Using (19) and definition (21), one has:

$$\tilde{\mathbf{w}}(k+1) = \tilde{\mathbf{w}}(k) - \beta\frac{\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2}f[e(k)]. \tag{22}$$

The left multiplication of both sides of (22) by $\mathbf{x}^T(k)$ leads to:

$$\overline{e}_{\mathrm{p}}(k) = e_{\mathrm{a}}(k) - \beta f[e(k)], \tag{23}$$

where the following error measures are considered:

$$\overline{e}_{\mathrm{p}}(k) \triangleq \mathbf{x}^T(k)\tilde{\mathbf{w}}(k+1), \tag{24}$$
$$e_{\mathrm{a}}(k) \triangleq \mathbf{x}^T(k)\tilde{\mathbf{w}}(k). \tag{25}$$

The application of (23) in (22) reveals that

$$\tilde{\mathbf{w}}(k+1) + \frac{\mathbf{x}(k)e_{\mathrm{a}}(k)}{\|\mathbf{x}(k)\|^2} = \tilde{\mathbf{w}}(k) + \frac{\mathbf{x}(k)\overline{e}_{\mathrm{p}}(k)}{\|\mathbf{x}(k)\|^2}, \tag{26}$$

which implies:

$$\left[\tilde{\mathbf{w}}^T(k+1)+\frac{\mathbf{x}^T(k)e_\mathrm{a}(k)}{\|\mathbf{x}(k)\|^2}\right]\left[\tilde{\mathbf{w}}(k+1)+\frac{\mathbf{x}(k)e_\mathrm{a}(k)}{\|\mathbf{x}(k)\|^2}\right] = \left[\tilde{\mathbf{w}}^T(k)+\frac{\mathbf{x}^T(k)\bar{e}_\mathrm{p}(k)}{\|\mathbf{x}(k)\|^2}\right]\left[\tilde{\mathbf{w}}(k)+\frac{\mathbf{x}(k)\bar{e}_\mathrm{p}(k)}{\|\mathbf{x}(k)\|^2}\right]. \quad (27)$$

After some manipulations, one obtains:

$$\|\tilde{\mathbf{w}}(k+1)\|^2+\frac{e_\mathrm{a}^2(k)}{\|\mathbf{x}(k)\|^2}=\left\|\tilde{\mathbf{w}}(k)\right\|^2+\frac{\bar{e}_\mathrm{p}(k)}{\|\mathbf{x}(k)\|^2}, \quad (28)$$

which provides a description *without approximations* of the energy flow through each iteration of the advanced algorithm. Note that (28) holds even in the case of a colored measurement noise. The groundwork for (28) is based on a basic energy conservation relationship that was originally established in [37], while dealing with the robustness analysis of adaptive filters.

The application of the expectation operator $\mathbb{E}[\cdot]$ in both sides of (28) results in

$$\mathbb{E}\left[\|\tilde{\mathbf{w}}(k+1)\|^2\right]+\mathbb{E}\left[\frac{e_\mathrm{a}^2(k)}{\|\mathbf{x}(k)\|^2}\right] = \mathbb{E}\left[\|\tilde{\mathbf{w}}(k)\|^2\right]+\mathbb{E}\left[\frac{\bar{e}_\mathrm{p}^2(k)}{\|\mathbf{x}(k)\|^2}\right]. \quad (29)$$

Assuming that the algorithm operates in steady-state under a stable condition enables one to derive that $\mathbb{E}\left[\|\tilde{\mathbf{w}}(k+1)\|^2\right]=\mathbb{E}\left[\|\tilde{\mathbf{w}}(k)\|^2\right]$. Thus:

$$\mathbb{E}\left[\frac{e_\mathrm{a}^2(k)}{\|\mathbf{x}(k)\|^2}\right]=\mathbb{E}\left[\frac{\bar{e}_\mathrm{p}^2(k)}{\|\mathbf{x}(k)\|^2}\right]. \quad (30)$$

When the filter presents a large number of taps and considering that the denominator in (30) is equal to $\|\mathbf{x}(k)\|^2$, the expected value of the ratio can be approximated by the ratio of the expected values. This is a reasonable approximation, since at the steady state the errors $e_\mathrm{a}(k)$ and $\bar{e}_\mathrm{p}(k)$ exhibit low sensitivity to the input data [38]. Therefore, the identity (30) collapses to

$$\mathbb{E}\left[e_\mathrm{a}^2(k)\right]\approx\mathbb{E}\left[\bar{e}_\mathrm{p}^2(k)\right]. \quad (31)$$

The derivation of a closed-form that predicts the steady-state MSE requires some additional steps. Since the errors $\{\bar{e}_\mathrm{p}(k), e_\mathrm{a}(k), e(k)\}$ are related by (see (1) and (5)):

$$\bar{e}_\mathrm{p}(k)=e_\mathrm{a}(k)-\beta f[e(k)], \quad (32)$$

one has

$$\bar{e}_\mathrm{p}^2(k)=e_\mathrm{a}^2(k)-2\beta e_\mathrm{a}(k)f[e(k)]+\beta^2 f^2[e(k)]. \quad (33)$$

After the application of (33) in (31), one arrives at:

$$2\mathbb{E}\{e_\mathrm{a}(k)f[e(k)]\}=\beta\mathbb{E}\left\{f^2[e(k)]\right\}. \quad (34)$$

Thus, it is assumed that $e_\mathrm{a}(k)$ and $e(k)$ are jointly Gaussian (a reasonable assumption for long adaptive filters [35]). Let one consider

$$\xi \triangleq \mathbb{E}\left[e^2(k)\right] \quad (35)$$

and

$$\rho \triangleq e_\mathrm{a}(k)e(k). \quad (36)$$

Employing the Price Theorem [39], one is led to:

$$\frac{\partial\mathbb{E}[e_\mathrm{a}(k)f[e(k)]]}{\partial\rho} = \mathbb{E}\left[\frac{\partial e_\mathrm{a}(k)}{\partial e_\mathrm{a}(k)}\frac{\partial f[e(k)]}{\partial e(k)}\right] = \int_{-\infty}^{\infty}\frac{[\delta[e(k)-\overline{\gamma}]+\delta[e(k)+\overline{\gamma}]}{\sqrt{2\pi\xi}}\exp\left[-\frac{e^2(k)}{2\xi}\right]de(k) = \frac{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]+\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}{\sqrt{2\pi\xi}}=\frac{2\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}{\sqrt{2\pi\xi}}, \quad (37)$$

which takes one to:

$$\mathbb{E}\{e_\mathrm{a}(k)f[e(k)]\}=\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\mathbb{E}[e_\mathrm{a}(k)e(k)]. \quad (38)$$

Assuming that the error is distributed according to a Gaussian distribution results in

$$\mathbb{E}\left\{f^2[e(k)]\right\}=2\beta\int_{\overline{\gamma}}^{\infty}\frac{1}{\sqrt{2\pi\xi}}\exp\left[-\frac{e^2(k)}{2\xi}\right]de(k) \quad (39)$$
$$=\frac{2}{\sqrt{\pi}}\int_{\frac{\overline{\gamma}}{\sqrt{\pi}}}^{\infty}\exp\left(-u^2\right)du=\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right),$$

where

$$\operatorname{erfc}(x)\triangleq\frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp(-t^2)dt. \quad (40)$$

The utilization of equations (38), (39) and (34) brings one to:

$$2\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\xi-\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\sigma_\nu^2 = \beta\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right), \quad (41)$$

where σ_ν^2 denotes the variance of the measurement noise $\nu(k)$, supposed to be zero-mean, independent, identically distributed and independent of the input signal. From (41) one arrives to the following *fixed-point* equation for the asymptotic mean square error:

$$\xi=\sigma_\nu^2+\frac{\beta}{2}\frac{\sqrt{\frac{\pi\xi}{2}}\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}. \quad (42)$$

Remark: Equation (42) is the main contribution of this section. It is a nonlinear equation whose positive solution is a theoretical approximation of the steady-state MSE of the SM-sign-NLMS algorithm. Interestingly, Equation (42) implies that the statistics of the input signal (e.g., its variance) do not impact the asymptotic performance of the algorithm.

IV. TRANSIENT ANALYSIS

In order to cope with the nonlinear characteristics of the update equation of the SM-sign-NLMS algorithm, a simple model for the input signal is adopted such as to evaluate its convergence analysis. The model originates a description of the modal behavior of the second-order behavior of the algorithm. The resulting description permits one to extract novel perceptions about the eigenvalue-distribution-dependent convergence behavior of the SM-sign-NLMS algorithm.

The model utilizes the ubiquitous independence assumption [40], which implies that vectors $\mathbf{x}(k)$ are independent and identically distributed. The eigendecomposition of the input autocovariance matrix $\mathbf{R}$

$$\mathbf{R} \triangleq \mathbb{E}\left[\mathbf{x}(k)\mathbf{x}^T(k)\right] = \mathbf{V}\Lambda\mathbf{V}^T = \sum_{i=0}^{N-1} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \tag{43}$$

emphasizes two import quantities for the following analysis: the eigenvalues $\lambda_i \in \mathbb{R}+$ (for $i \in \{0, 1, \ldots, N-1\}$) and the orthonormal eigenvectors $\mathbf{v}_i \in \mathbb{R}^N$ (for $i \in \{0, 1, \ldots, N-1\}$).

A simple model that fits the first- and second-order moments of $\mathbf{x}(k)$ is [41]:

$$\mathbf{x}(k) = s_k r_k \tilde{\mathbf{v}}(k), \tag{44}$$

where *statistically-independent* random variables[2] s_k (signal), r_k (*radial* distribution) and $\tilde{\mathbf{v}}(k)$ (discrete *angular* distribution) are distributed according to

$$\text{Prob}\{s_k = \pm 1\} = \frac{1}{2}, \tag{45}$$

$$r(k) \sim \|\mathbf{x}(k)\|, \tag{46}$$

$$\text{Prob}\{\tilde{\mathbf{v}}(k) = \mathbf{v}_i\} = \frac{\lambda_i}{\text{Tr}[\mathbf{R}]}, \tag{47}$$

where (46) means that $r(k)$ emulates the distribution of the original (*i.e.*, obtained in a tapped-delay structure) $\mathbf{x}(k)$. Note that the Gaussianity of $\mathbf{x}(k)$ can be partially inserted into the model, which motivates the *Price heuristic* [16] that will be adopted.

In order to advance a second-order model for the SM-sign-NLMS algorithm, consider the multiplication of both sides of (22) by their transpose. Using the expectation operator, one arrives at the following recursion:

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{w}}}(k+1) = \mathbf{R}_{\tilde{\mathbf{w}}}(k) &- \beta\mathbb{E}\left\{\frac{\tilde{\mathbf{w}}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\} \\ &- \beta\mathbb{E}\left\{\frac{\mathbf{x}(k)\tilde{\mathbf{w}}^T(k)}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\} \\ &+ \beta^2\mathbb{E}\left\{\frac{\mathbf{x}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^4} f^2[e(k)]\right\},\end{aligned} \tag{48}$$

where

$$\mathbf{R}_{\tilde{\mathbf{w}}}(k) \triangleq \mathbb{E}\left[\tilde{\mathbf{w}}(k)\tilde{\mathbf{w}}^T(k)\right] \tag{49}$$

is the weight-error autocorrelation matrix. This matrix plays a fundamental role for the prediction of the MSE, because it permits the computation of

$$\tilde{\lambda}_i(k) \triangleq \mathbf{v}_i^T \mathbf{R}_{\tilde{\mathbf{w}}}(k)\mathbf{v}_i, \tag{50}$$

from which, with the independence assumption and the common assumptions for the measurement noise, the MSE can be obtained:

$$\begin{aligned}\xi(k) \triangleq \mathbb{E}\left[e^2(k)\right] &= \sigma_v^2 + \mathbb{E}\left[\|\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\|^2\right] \\ &= \sigma_v^2 + \text{Tr}\left[\Sigma\mathbf{V}^T\mathbf{R}_{\tilde{\mathbf{w}}}\mathbf{V}\right] \\ &= \sigma_v^2 + \sum_{i=0}^{N-1}\lambda_i\tilde{\lambda}_i(k).\end{aligned} \tag{51}$$

Equation (51) implies that predicting the dynamics of $\tilde{\lambda}_i(k)$ allows one to estimate the evolution of the MSE. Thus, multiplying (48) on the left by $\mathbf{v}_i^T$ and on the right by $\mathbf{v}_i$, one has:

$$\begin{aligned}\tilde{\lambda}_i(k+1) = \tilde{\lambda}_i(k) &\underbrace{- \beta\mathbb{E}\left\{\frac{\mathbf{v}_i^T\tilde{\mathbf{w}}(k)\mathbf{x}^T(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\}}_{I} \\ &\underbrace{- \beta\mathbb{E}\left\{\frac{\mathbf{v}_i^T\mathbf{x}(k)\tilde{\mathbf{w}}^T(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\}}_{II} \\ &\underbrace{+ \beta^2\mathbb{E}\left\{\frac{\mathbf{v}_i^T\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^4} f^2[e(k)]\right\}}_{III}.\end{aligned} \tag{52}$$

The analytic computation of terms I-III is a challenging task, especially due to the presence of the nonlinear terms. In order to circumvent such an issue, two simplifying assumptions are adopted: the model (45)-(47) and the Price heuristic (discussed in [16]). Briefly, the Price heuristic utilizes the Price theorem [42] (which is strictly valid only when jointly Gaussian random variables are involved in the expectation) when the input vector is generated according to the model described by (45)-(47). Therefore, term I in (52) can be expressed as:

$$\begin{aligned}I &= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\bar{\gamma}^2}{2\xi}\right]\mathbb{E}\left[\frac{\mathbf{v}_i^T\tilde{\mathbf{w}}(k)\mathbf{x}^T(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^2}\right] \\ &= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\bar{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\text{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k).\end{aligned} \tag{53}$$

whereas terms II and III can be simplified to

$$\begin{aligned}II &= -\beta\mathbb{E}\left[\frac{\mathbf{v}_i^T\mathbf{x}(k)\tilde{\mathbf{w}}^T(k)f[e(k)]\mathbf{v}_i}{\|\mathbf{x}(k)\|^2}\right] \\ &= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\bar{\gamma}^2}{2\xi}\right]\mathbb{E}\left[\frac{\mathbf{v}_i^T\mathbf{x}(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}^T(k)\tilde{\mathbf{w}}(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^2}\right]\end{aligned}$$

[2] In the case of $\mathbf{v}(k)$, the adequate expression is "random vector".

$$= -\beta\sqrt{\frac{2}{\pi\xi}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\mathrm{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k) \quad (54)$$

$$,III \approx \beta^2\mathbb{E}\left[\frac{\mathbf{v}_i^T\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{v}_i}{\|\mathbf{x}(k)\|^4}\right]\mathbb{E}\left\{f^2[e(k)]\right\}$$

$$= \beta^2\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)\sum_{j=0}^{N-1}\frac{\lambda_j}{\mathrm{Tr}[\mathbf{R}]}\frac{\mathbb{E}[s_k^2]\mathbb{E}\left[r_k^2\right]\mathbf{v}_i^T\mathbf{v}_j\mathbf{v}_j^T\mathbf{v}_i}{\mathbb{E}\left[s_k^4\right]\mathbb{E}\left[r_k^4\right]\mathbf{v}_j^T\mathbf{v}_j}$$

$$= \beta^2\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\pi}}\right)\frac{\lambda_i}{\mathbb{E}[r^4]} \quad (55)$$

The combination (52)- (55) of yields the subsequent recursion:

$$\tilde{\lambda}_i(k+1) = \tilde{\lambda}_i(k) - 2\sqrt{\frac{2}{\pi\xi}}\beta\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]\frac{\lambda_i}{\mathrm{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k) + \beta^2\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)\frac{\lambda_i}{\mathbb{E}[r^4]}, \quad (56)$$

where one observes that the dynamics of the modes are coupled, since ξ depends on all $\tilde{\lambda}_i(k)$ (see (51)). Using (56) and (51), it is possible to predict the evolution of the MSE. Therefore, a model for the transient behavior of the algorithm was obtained.

Such a model is also valid in asymptotic regime. In this case, assuming that the algorithm operates in a stable manner, one has that $\tilde{\lambda}_i(k+1) = \tilde{\lambda}_i(k)$ and

$$\tilde{\lambda}_i(k) = \frac{\beta}{2}\sqrt{\frac{\pi\xi}{2}}\frac{\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\frac{\mathrm{Tr}[\mathbf{R}]}{\mathbb{E}\left[r^4\right]}. \quad (57)$$

Which implies that in steady-state:

$$\xi = \sigma_v^2 + \sum_{i=0}^{N-1}\lambda_i\tilde{\lambda}_i(k) = \sigma_v^2 + \frac{\beta}{2}\frac{\sqrt{\frac{\pi\xi}{2}}\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\exp\left(-\frac{\overline{\gamma}^2}{2\xi}\right)}\frac{\{\mathrm{Tr}[\mathbf{R}]\}^2}{\mathbb{E}\left[r^4\right]}, \quad (58)$$

an identity similar to the one obtained with energy-conservation arguments (see (42)).

Remark: It is possible to elucidate under which conditions Equations (58) and (42) are equivalent. Assume that the input vector derives from a white Gaussian process. Hence,

$$\mathbb{E}\left[\|\mathbf{x}(k)\|_2^n\right] = 2^{\frac{n-2}{2}}\sigma_x^n N\frac{\Gamma\left(\frac{N+n}{2}\right)}{\Gamma\left(\frac{N+2}{2}\right)}, \quad (59)$$

where σ_x^2 is the variance of x and

$$\Gamma(x) \triangleq \int_0^{\infty} t^{x-1}\exp(-t)dt. \quad (60)$$

Thus,

$$\frac{\{\mathrm{Tr}[\mathbf{R}]\}^2}{\mathbb{E}\left[r^4\right]} = \frac{N}{2}\frac{\Gamma\left(\frac{N+2}{2}\right)}{\Gamma\left(\frac{N+4}{2}\right)}. \quad (61)$$

Since

$$\lim_{N\to\infty}\frac{N}{2}\frac{\Gamma\left(\frac{N+2}{2}\right)}{\Gamma\left(\frac{N+4}{2}\right)} \approx 1, \quad (62)$$

both equations are equivalent for large values of N.

V. DESIGN OF A VARIABLE STEP-SIZE SM-SIGN-NLMS

This section addresses the issue of deriving a time-varying step size in order to tackle the compromise between low asymptotic MSE and fast convergence rate. The obtained step-size sequences agrees with intuition and are obtained based on the theoretical model described in Section IV.

Consider a time-variant step size $\beta(k)$ in (56). The minimization of $\tilde{\lambda}_i(k+1)$ can be enforced by

$$\frac{\partial\tilde{\lambda}_i(k+1)}{\partial\beta(k)} = -2\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\lambda_i}{\mathrm{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k) + 2\beta(k)\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)\frac{\lambda_i}{\mathbb{E}\left[r^4\right]} = 0. \quad (63)$$

Therefore, in order to maximize the convergence rate of the algorithm, the following theoretically-based choice can be made:

$$\beta(k) = \frac{\mathbb{E}[r^4]}{\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)}\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\tilde{\lambda}_i(k)}{\mathrm{Tr}[\mathbf{R}]}. \quad (64)$$

Unfortunately, (64) is not a feasible choice in practice, because $\tilde{\lambda}_i$ and $\mathrm{Tr}[\mathbf{R}]$ are not observable. Hence, some approximations should be performed. Assuming a white input signal, one may write:

$$\tilde{\lambda}_i(k) \approx \sigma_x^2\sigma_{\tilde{w}}^2 \approx \frac{\xi - \sigma_v^2}{N}, \quad (65)$$

where $\sigma_{\tilde{w}}^2$ is the variance of the deviations $\tilde{w}_i^2(k)$, supposed to be constant along the taps. Combining (65) with the approximation $\mathbb{E}[r^4] \approx N^2\sigma_x^2$ leads to

$$\beta_{\mathrm{opt}}(k) = \sqrt{\frac{2}{\pi\xi(k)}}\frac{\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]}{\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)}\left[\xi(k) - \hat{\sigma}_v^2\right], \quad (66)$$

where $\hat{\sigma}_v^2$ is an estimate of σ_v^2. Note that σ_v^2 is already estimated in the Set-membership approach (see (6)) and $\xi(k)$ can be estimated in an online manner through

$$\hat{\xi}(k+1) = \overline{\lambda}\hat{\xi}(k) + (1-\overline{\lambda})e^2(k), \quad (67)$$

where

$$\overline{\lambda} = 1 - \frac{1}{KN}, \quad (68)$$

with $K \geq 2$. In fact, the resulting variable step size is almost insensitive to the value of K, so that one can impose $K = 2$ in order to avoid the adjustment of an additional parameter [17].

VI. TRACKING

The ability to operate in a nonstationary setting is one of the most desirable features of adaptive filtering algorithms [40]. The learning behavior of an adaptive filter is more sophisticated in this case, since time-varying plants cause a "lag" in the adaptive learning process. Thus, the tracking capabilities of an algorithm are enhanced when the step size has large values, whereas the variance of the adaptive estimator diminishes under small values of β. This implies that asymptotic performance is optimized when this trade-off is taken into account.

A time-variant feature of the ideal plant is explicitly introduced by the following first-order stochastic multivariate Markovian random walk model:

$$\mathbf{w}^\star(k+1) = \mathbf{w}^\star(k) + \mathbf{q}(k), \tag{69}$$

where $\mathbf{q}(k)$ is a zero-mean random vector that is statistically independent from the remaining random variables. It also has the following covariance matrix

$$\mathbf{Q} = \mathbb{E}\left[\mathbf{q}(k)\mathbf{q}^T(k)\right] = \sigma_q^2 \mathbf{I}_N. \tag{70}$$

Remark: The random walk model is popular in the area of adaptive filtering. Its suitability is discussed in more detail in [43]. It should be noted that first-order perturbations (69) are encountered in applications such as acoustic echo cancellation and transmission systems [40], in such a manner that its use is not due solely to analytical convenience.

Under (69), a new formulation is called for Equation (22):

$$\tilde{\mathbf{w}}(k+1) = \tilde{\mathbf{w}}(k) - \beta \frac{\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2} f[e(k)] + \mathbf{q}(k). \tag{71}$$

In a similar manner, Equation (48) can be restated in the form of:

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{w}}}(k+1) = \mathbf{R}_{\tilde{\mathbf{w}}}(k) &- \beta\mathbb{E}\left\{\frac{\tilde{\mathbf{w}}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\} + \mathbf{Q} \\ &- \beta\mathbb{E}\left\{\frac{\mathbf{x}(k)\tilde{\mathbf{w}}^T(k)}{\|\mathbf{x}(k)\|^2} f[e(k)]\right\} \\ &+ \beta^2\mathbb{E}\left\{\frac{\mathbf{x}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^4} f^2[e(k)]\right\}.\end{aligned} \tag{72}$$

Thus, a revision is needed for equation (73) to read:

$$\begin{aligned}\tilde{\lambda}_i(k+1) = \tilde{\lambda}_i(k) &+ \beta^2(k)\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right)\frac{\lambda_i}{\mathbb{E}\left[r^4\right]} + \sigma_q^2 \\ &- 2\sqrt{\frac{2}{\pi\xi(k)}}\beta(k)\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\lambda_i}{\mathrm{Tr}[\mathbf{R}]}\tilde{\lambda}_i(k).\end{aligned} \tag{73}$$

Through various mathematical manipulations (akin to those that gave rise to Equation (58)), one comes to the subsequent fixed-point equation describing the asymptotic behavior of the algorithm in the non-stationary setting:

$$\begin{aligned}\xi = \sigma_\nu^2 &+ \frac{\beta\sqrt{\pi\xi}}{2\sqrt{2}}\frac{\mathrm{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi}}\right)}{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\frac{\{\mathrm{Tr}[\mathbf{R}]\}^2}{\mathbb{E}\left[r^4\right]} \\ &+ \frac{N\sqrt{\pi\xi}}{2\sqrt{2}\beta}\frac{\mathrm{Tr}[\mathbf{R}]}{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\sigma_q^2,\end{aligned} \tag{74}$$

where the novel additive term $\frac{N\sqrt{\pi\xi}}{2\sqrt{2}\beta}\frac{\mathrm{Tr}[\mathbf{R}]}{\exp\left[-\frac{\overline{\gamma}^2}{2\xi}\right]}\sigma_q^2$ contains the analytical contribution of the Markovian perturbation $\mathbf{q}(k)$ to the asymptotic performance of the algorithm.

VII. DEFICIENT-LENGTH ANALYSIS

In practice, the length of the ideal plant is greater than the adaptive filter length. Sometimes, the designer chooses such a configuration in order to deal with computational limitations [44] or when the convergence rate should be increased [45]. Thus, it is important to characterize the algorithm performance under suboptimal operation [46]. In the following, a theoretical approach distinct from the previous ones is described. The stochastic modeling is divided in two parts: *i)* first-order analysis; and *ii)* second-order analysis.

A. FIRST-ORDER BEHAVIOR

In the deficient-length configuration, the desired signal can be described as

$$d(k) = \left[\mathbf{w}^\star\right]^T \mathbf{x}(k) + \left[\overline{\mathbf{w}}^\star\right]^T \overline{\mathbf{x}}(k) + \nu(k), \tag{75}$$

where $\overline{\mathbf{w}}^\star \in \mathbb{R}^L$ and

$$\overline{\mathbf{x}}(k) \triangleq \left[x(k-N) x(k-N-1) \ldots x(k-N-L+1)\right]^T, \tag{76}$$

where the ideal vector has length $N+L$ (with N denoting the length of the adaptive filter, as before). In this setting, recursion (22) is still valid when the error is written as

$$e(k) = \tilde{\mathbf{w}}^T(k)\mathbf{x}(k) + \left[\overline{\mathbf{w}}^\star\right]^T \overline{\mathbf{x}}(k) + \nu(k). \tag{77}$$

Thus, using (22) and (77), applying the expectation operator and the Prize heuristic alongside the independence noise assumption leads to:

$$\mathbb{E}\left[\tilde{\mathbf{w}}(k+1)\right] = \mathbb{E}\left[\tilde{\mathbf{w}}(k)\right] - \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\mathcal{A} \tag{78}$$

where

$$\mathcal{A} \triangleq \mathbb{E}\left\{\frac{\mathbf{x}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2}\tilde{\mathbf{w}}(k) + \frac{\overline{\mathbf{x}}^T(k)\overline{\mathbf{w}}^\star\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2} + \frac{\mathbf{x}(k)\nu(k)}{\|\mathbf{x}(k)\|^2}\right\}. \tag{79}$$

Using the approximation $\mathbb{E}\left[\frac{1}{\|\mathbf{x}(k)\|^2}\right] \approx \frac{1}{N\sigma_x^2}$ yields:

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{w}}(k+1)\right] &= \mathcal{B}\mathbb{E}\left[\tilde{\mathbf{w}}(k)\right] \\ &- \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left(-\frac{\overline{\gamma}^2}{2\xi(k)}\right)\vartheta,\end{aligned} \tag{80}$$

where

$$\mathcal{B} \triangleq \left(\mathbf{I} - \beta\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\frac{\mathbf{R}}{N\sigma_x^2}\right), \tag{81}$$

and

$$\vartheta \triangleq \frac{\mathbb{E}\left[\overline{\mathbf{x}}^T(k)\overline{\mathbf{w}}^\star \mathbf{x}(k)\right]}{N\sigma_x^2}. \tag{82}$$

Assuming a white input, Equation (81) implies that the SM-sign-NLMS algorithm is unbiased, in the sense that the adaptive weight vector converges in the mean to the first N elements of the ideal plant:

$$\mathbb{E}\left[\tilde{\mathbf{w}}(\infty)\right]_{\text{white}} = \mathbf{0} \Rightarrow \mathbb{E}\left[\mathbf{w}(\infty)\right] = \mathbf{w}^\star. \tag{83}$$

B. SECOND-ORDER ANALYSIS

The second-order behavior of the algorithm is more involved. Using (77) leads to:

$$\begin{aligned} e^2(k) = {} & \tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\tilde{\mathbf{w}}(k) + 2\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\overline{\mathbf{x}}^T(k)\overline{\mathbf{w}}^\star \\ & + \left[\overline{\mathbf{w}}^\star\right]^T \overline{\mathbf{x}}(k)\overline{\mathbf{x}}^T(k)\overline{\mathbf{w}}^\star + \nu^2(k) + \mathcal{O}[\nu(k)]. \end{aligned} \tag{84}$$

After the application of the expectation operator and the usual statistical assumptions, one has:

$$\begin{aligned} \mathbb{E}\left[e^2(k)\right] = {} & \sigma_\nu^2 + \left[\overline{\mathbf{w}}^\star\right]^T \mathbf{R}_{N-M}\overline{\mathbf{w}}^\star + 2\mathbf{b}^T\mathbb{E}\left[\tilde{\mathbf{w}}(k)\right] \\ & + \operatorname{Tr}\left\{\mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}}(k)\right\}, \end{aligned} \tag{85}$$

where

$$\mathbf{b} \triangleq \mathbb{E}\left\{\left[\overline{\mathbf{x}}^T(k)\overline{\mathbf{w}}^\star\right]\mathbf{x}(k)\right\}, \tag{86}$$

$$\mathbf{R}_{N-M} \triangleq \mathbb{E}\left[\overline{\mathbf{x}}(k)\overline{\mathbf{x}}^T(k)\right]. \tag{87}$$

Multiplying (22) by its transpose and applying the expectation operator with the usual simplifying assumptions leads to, after some mathematical manipulations:

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{w}}}(k+1) = {} & \mathbf{R}_{\tilde{\mathbf{w}}}(k) - g\mathbf{R}_{\tilde{\mathbf{w}}}(k)\mathbf{R} - g\mathbf{R}\mathbf{R}_{\tilde{\mathbf{w}}}(k) \\ & - g\mathbb{E}\left[\tilde{\mathbf{w}}(k)\right]\mathbb{E}\left\{\mathbf{x}^T(k)\left\{\left[\overline{\mathbf{w}}^\star\right]^T\overline{\mathbf{x}}(k)\right\}\right\} \\ & + \frac{\beta^2}{\mathbb{E}\left[r^4\right]}\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \end{aligned} \tag{88}$$

where

$$g \triangleq \frac{\beta}{\mathbb{E}\left[\|\mathbf{x}(k)\|^2\right]}\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]. \tag{89}$$

Assuming a white input signal, utilizing the approximation $\mathbb{E}\left[\|\mathbf{x}(k)\|^2\right] = N\sigma_x^2$ and applying the trace operator in (88) yields:

$$\begin{aligned} \theta(k+1) = {} & \theta(k) - \frac{2\beta}{N}\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\theta(k) \\ & + \frac{N\beta^2}{\mathbb{E}\left[r^4\right]}\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \end{aligned} \tag{90}$$

where

$$\theta(k) \triangleq \operatorname{Tr}\left[\mathbf{R}_{\tilde{\mathbf{w}}}(k)\right], \tag{91}$$

is the mean square deviation (MSD) of the algorithm.

The assumption of a white Gaussian input signal simplifies the recursion to:

$$\begin{aligned} \theta(k+1) = {} & \theta(k) - \frac{2\beta}{N}\sqrt{\frac{2}{\pi\xi(k)}}\exp\left[-\frac{\overline{\gamma}^2}{2\xi(k)}\right]\theta(k) \\ & + \frac{\beta^2\Gamma\left(\frac{N+2}{2}\right)}{2\sigma_x^4\Gamma\left(\frac{N+4}{2}\right)}\operatorname{erfc}\left(\frac{\overline{\gamma}}{\sqrt{2\xi(k)}}\right), \end{aligned} \tag{92}$$

where $\theta(k)$ and $\xi(k)$ are related by:

$$\xi(k) = \sigma_\nu^2 + \sigma_x^2\|\overline{\mathbf{w}}^\star\|^2 + \sigma_x^2\theta(k). \tag{93}$$

Equations (92) and (93) specify a coupled pair of fixed-point equations. This implies that the asymptotic MSD and MSE can be estimated through an iterative refinement of an initial guess for $\theta(k)$. At each iteration, the current estimation of the MSD $\theta(k)$ leads to a novel estimate of the MSE $\xi(k)$ through (93) and the current estimates of $\xi(k)$ and $\theta(k)$ can be utilized for updating the estimate of $\theta(k)$ through (92). Thus, Equations (92) and (93) provides theoretical estimates for both asymptotic MSD and MSE of the advanced SM-sign-NLMS algorithm. Note that g presents a strong stochastic coupling with the random vectors $\mathbb{E}\left[\tilde{\mathbf{w}}(k)\right]$ and $\mathbb{E}\left\{\mathbf{x}^T(k)\left\{\left[\overline{\mathbf{w}}^\star\right]^T\overline{\mathbf{x}}(k)\right\}\right\}$, which, by simplicity, was neglected in the stochastic analysis. This independence assumption unfortunately restricts the analysis for configurations where the step size is small.

VIII. RESULTS

In the absence of explicit instructions to the contrary, the number of independent Monte Carlo trials was set to $K' = 1000$ for all experiments. Three distinct input signals are adopted in the different scenarios considered. The first of them is sampled from an unit-variance zero-mean white Gaussian process (UZWG). Other alternatives are a fourth-order autoregressive (AR(4)) process computed through $x(k) = 0.75x(k-1) + 0.19x(k-2) + 0.09x(k-3) - 0.5x(k-4) + g(k)$, where $g(k)$ is a zero-mean white Gaussian noise with unit variance (see [47]) or a MA(2) process, where the input signal is obtained by filtering an UZWG by the filter $1 - 0.8z^{-1} + 0.2z^{-2}$. The unknown plants to be identified are the ones measured and described in [48]. Unless stated to the contrary, the noise signal $\nu(k)$ is zero mean, white and Gaussian. In all experiments, the adaptive filter is initialized with zeros, and its length is the same than the one of the ideal plant (except in the deficient-length configuration).

A. COMPARISON WITH THE STANDARD SM-NLMS ALGORITHM

The advantage of the advanced algorithm in an environment subject to impulsive noise can be observed in a very simple setup. Consider the unknown plant to be the sixth one of [48] ($N = 120$). Additionally, $\tau = 5$, $\beta_{\text{SM-sign}} = 0.01$ and $\beta_{\text{SM}} = 0.03$. Learning coefficients were picked to ensure the same asymptotic performance of the algorithms. The

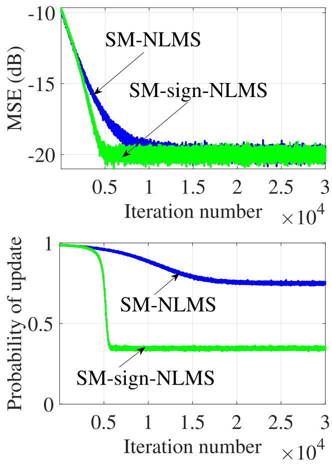


FIGURE 2. **Evolution of the MSE and probability of update of both SM-sign-NLMS and SM-NLMS algorithms. (a) MSE; (b) Probability of update. The curves were computed through 10^4 independent Monte Carlo trials.**

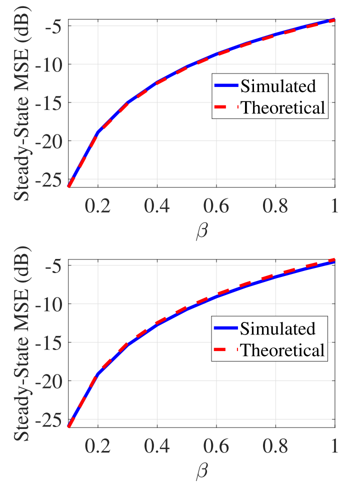


FIGURE 3. **Steady-state MSE (dB) of the advanced algorithm. (a) White input signal; (b) AR(4) input signal.**

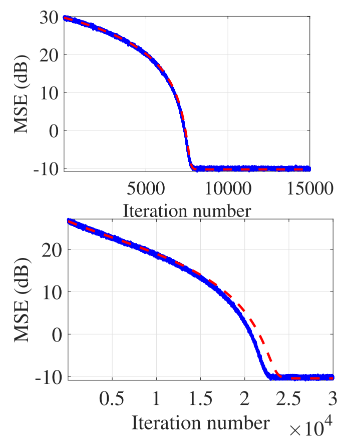


FIGURE 4. **Transient behavior of the proposed algorithm. (a) White input signal; (b) MA(3) input signal. Solid blue: simulated. Dashed red: theoretical prediction.**

variance of the additive noise is equal to $\sigma_\nu^2 = 10^{-6}$. Besides the measurement noise $\nu(k)$, an impulsive noise $\eta(k)$ is introduced into the reference signal $d(k)$. The impulsive noise is created as $\eta(k) = \omega_k \mathcal{N}_k$, where ω_k follows a Bernoulli process with a success probability of $P[\omega_k = 1] = 0.1$, and $\mathcal{N}_k$ is a zero-mean Gaussian noise with a variance equal to $\sigma_\eta^2 = 0.1$. Figure 2 depicts the results. Note that the advanced algorithm presents better rate convergence than the original SM-NLMS algorithm. Furthermore, it presents a much lower asymptotic probability of update (34.49%), less than half than the one obtained by the SM-NLMS algorithm (74.87%). This indicates that the proposed algorithm shows great potential in terms of reducing computational burden, as well as in its data censoring abilities.

B. STEADY-STATE BEHAVIOR

To assess the stochastic model against experimental data, choose a configuration involving the first plant of [48] ($N = 64$), with $\overline{\gamma} = 5$ and $\sigma_\nu^2 = 10^{-5}$. The fixed-point equation (42) was iterated from an initial value $\xi_0 = 10^{-2}$ until the absolute difference between two consecutive estimates of the MSE is less than 10^{-10}. Based on Figure 3, one can infer that the theoretical model demonstrates strong agreement with the data, for both white and colored input signals. The largest deviation between the theoretical prediction and simulated results is observed at $\beta = 1$, with a discrepancy of 0.0961 dB (resp. 0.2862 dB) for white (resp. colored) input signals. The increased discrepancy between theoretical predictions and simulated data at higher β values is attributed to the fact that, in this case, $e_a(k)$ and $\overline{e}_p(k)$ exhibit a stronger stochastic coupling with respect to the input data, resulting in a less accurate approximation (31).

C. TRANSIENT BEHAVIOR

The ability of the devised stochastic model for predicting the transient behavior is assessed in a setup where the ideal plant is the second one of [48] ($N = 96$), with $\beta = 0.5$ and $\tau = 2$. The comparison between the actual MSE curve and the predicted one can be seen in Figure 4. One may observe that the proposed model fits well the simulated curve, except in the case of a colored input signal, where model has some disparity in the final part of the transient regime. Nevertheless, it is important to observe that the model remains highly accurate during the early iterations, as well as in the steady state. Colored input signals elevate the stochastic coupling among the adaptive coefficients, making it difficult for certain theoretical models to precisely capture the behavior of the learning algorithm.

D. VARIABLE STEP-SIZE SCHEME

As the SM-sign-NLMS algorithm is introduced in this paper, there are no prior time-varying learning rate techniques that can be applied to it. Therefore, we will assess the performance of the proposed VSS method with various choices of the learning rate parameter (assumed to be fixed

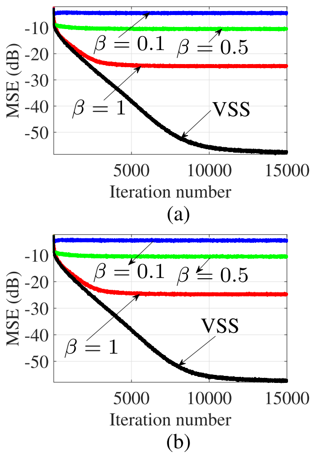


FIGURE 5. **Transient behavior of the VSS-SM-sign-NLMS algorithm, compared against the original version with a fixed step size (for $\beta \in \{0.1, 0.5, 1\}$). The input signal is AR(4). (a) estimated noise variance $\hat{\sigma}_\eta^2 = \sigma_\eta^2$; (b) estimated noise variance $\hat{\sigma}_\eta^2 = 2\sigma_\eta^2$.**

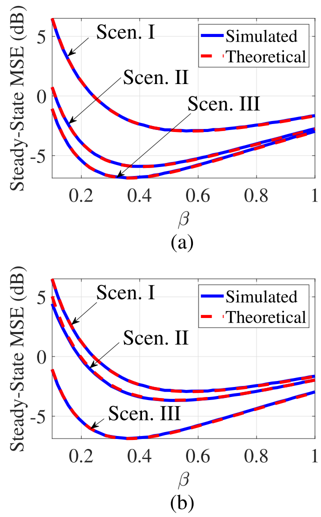


FIGURE 6. **Steady-state performance of the advanced algorithm for distinct values of β and three different scenarios. (a) White input signal; (b) MA(3) input signal.**

across iterations). It is known that small values of β favor asymptotic performance, while higher values accelerate the convergence rate. Consider $\tau = 3$, $\sigma_v^2 = 10^{-6}$ and the unknown plant being the third one of [48] (with $N = 96$). Figure 5 demonstrates that the VSS-based scheme outperforms the fixed step size version, with distinct values of β. The benefit persists even when the estimated variance of the noise variance is wrong (see Fig. 5.(b)). Thus, the VSS presents better asymptotic performance, without reduction in the convergence rate. The fact that the performance of the VSS version of the algorithm remains largely unchanged in the presence of a significant error in noise variance estimation is relevant, given that the process of estimating this variance is not straightforward.

E. TRACKING

Equation (74) provides a theoretical estimate of the steady-state performance of the SM-sign-NLMS algorithm. For the theoretical prediction, it was iterated from an initial value $\xi_0 = 10^{-2}$ until the absolute difference between two consecutive estimates of the MSE is less than 10^{-10}. In this section, its accuracy is assessed for both white and colored input signals. Three distinct scenarios are taken into account:

- *Scenario I*: $\tau = 5, \sigma_v^2 = 5 \times 10^{-6}, \sigma_q^2 = 2 \times 10^{-5}$;
- *Scenario II*: $\tau = 4, \sigma_v^2 = 10^{-6}, \sigma_q^2 = 10^{-5}$;
- *Scenario III*: $\tau = 3, \sigma_v^2 = 5 \times 10^{-7}, \sigma_q^2 = 8 \times 10^{-6}$.

The fourth plant of [48] (with $N = 128$) was adopted as the ideal system. Note that a high degree of non-stationarity was imposed, since σ_q^2 (the variance of the i.i.d. Gaussian Markovian perturbation) is higher than the usual in the literature (see, *e.g.*, [49], [50]). Figure 6 depicts the results. Note that the theoretical model predicts in an accurate manner the asymptotic performance of the algorithm, for both white and colored input signals.

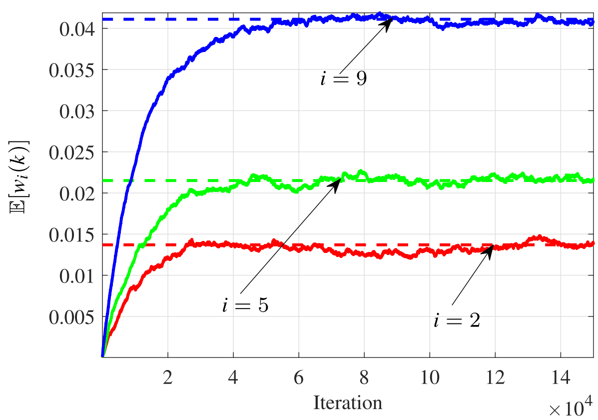


FIGURE 7. **Mean behavior of the weights $w_i(k)$ (for $i \in \{2, 5, 9\}$) of a deficient-length configuration of the sign-SM-NLMS algorithm. The input signal is white.**

F. FIRST-ORDER ANALYSIS OF THE DEFICIENT-LENGTH SCENARIO

Equation (83) implies that the proposed algorithm is asymptotically unbiased in the deficient-length case, in the sense that the coefficients of the adaptive filter converge to the first N coefficients of the optimal vector.[3] In order to quantitatively assess such a prediction, consider the case where the ideal solution contains the first 30 elements of the fifth plant of [48], whereas the adaptive filter operates with $N = 20$ adaptive taps. In this simulated scenario, $\tau = 2, \sigma_v^2 = 10^{-6}$, and $\beta = 5 \times 10^{-3}$. Figure 7 depicts the asymptotic convergence in the mean of three adaptive weights to the corresponding coefficient of the optimal solution, as expected by the devised theoretical model.

[3] Note that the derivation of (83) assumed a white input signal.

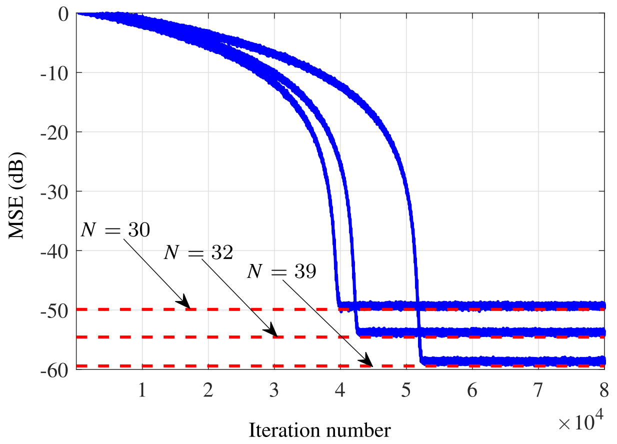


FIGURE 8. **Theoretical prediction (dashed red horizontal lines) *versus* the MSE along the iterations (in blue solid), for the deficient-length configuration. The input signal is white.**

G. SECOND-ORDER ANALYSIS OF THE DEFICIENT-LENGTH SCENARIO

The solution of the coupled pair of fixed-point Equations (92) and (93) provides a theoretical estimate of the asymptotic MSE of the deficient-length configuration. In order to assess the theoretical prediction, consider a setup with $\tau = 2$, $\sigma_v^2 = 10^{-6}$, $\beta = 10^{-3}$ and

$$w_i^{\star} = \exp(-0.2i)\cos(0.3i), \tag{94}$$

for $i \in \{0, \ldots, 39\}$. The quantities θ and ξ are initialized with the value 10^{-2}. They are iterated through Equations (92) and (93) until the absolute difference between two consecutive estimates of the MSE is lesser than 10^{-10}. Figure 8 allows one to conclude that the theoretical model described in Section VII-B is able to estimate the asymptotic performance of distinct suboptimal lengths of the adaptive filter. Note however that the theoretical prediction slightly overestimates the asymptotic performance of the algorithm. Indeed, the difference between the experimental steady-state performance of the algorithm and the theoretical prediction is 0.593 dB (resp. 0.8034 dB) when $N = 30$ (resp. $N = 32$).

IX. CONCLUSION

This paper advances a novel algorithm that combines the Set-membership filtering concept with a signed normalized adaptive filter. The update equation is derived from a minimum disturbance description. A theoretical fixed-point equation that estimates its asymptotic MSE was derived, based on energy-conservation arguments. A transient analysis of the algorithm was provided, utilizing a simplified model for the input vector, which divorces its radial distribution from its (discrete) angular distribution. The transient analysis was employed to derive an efficient and nonparametric variable step size scheme that combines a faster convergence rate with better steady-state performance. Such an analysis also was generalized in order to capture the asymptotic behavior of the algorithm when the ideal plant is nonstationary and subjected to a first-order Markovian perturbation. A novel stochastic model for the deficient-length configuration was described, that predicts the unbiasedness of the algorithm when the input signal is white. Further, the mean square behavior of the algorithm in the suboptimal-length setting was also predicted from a stochastic model. The findings were confirmed by extensive simulations.

REFERENCES

[1] L. Li, Y.-F. Pu, S. S. Bhattacharjee, and M. G. Christensen, "Widely linear complex-valued hyperbolic secant adaptive filtering algorithm and its performance analysis," *Signal Process.*, vol. 212, Nov. 2023, Art. no. 109146.

[2] Q. Liu, X. Huo, K.-Z. Liu, and H. Zhao, "A parallelized input matching LMS adaptive filter for the rejection of spatially cyclic disturbances," *IEEE Trans. Ind. Electron.*, vol. 70, no. 10, pp. 10536–10545, Oct. 2023.

[3] B. Zhu, B. Wang, B. Cai, Y. Zhu, P. Chao, and Z. Fang, "A variable step size least mean p-power adaptive filtering algorithm based on multi-moment error fusion," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 1, Jul. 2023.

[4] M. R. Petraglia, E. L. Marques, and D. B. Haddad, "Low-complexity affine projection subband algorithm for robust adaptive filtering in impulsive noise," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.

[5] R. L. Das and M. Narwaria, "Lorentzian based adaptive filters for impulsive noise environments," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 6, pp. 1529–1539, Jun. 2017.

[6] Md. Z. A. Bhotto, M. O. Ahmad, and M. N. S. Swamy, "Robust shrinkage affine-projection sign adaptive-filtering algorithms for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3349–3359, Jul. 2014.

[7] S. Werner, M. L. R. deCampos, and P. S. R. Diniz, "Partial-update NLMS algorithms with data-selective updating," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 938–949, Apr. 2004.

[8] L. Guo and Y.-F. Huang, "Frequency-domain set-membership filtering and its applications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1326–1338, Apr. 2007.

[9] M. Lee, S. Moon, and P. Park, "Improved combined step-size sign subband adaptive filter algorithms with variable mixing factors," in *Proc. 22nd Int. Conf. Control, Autom. Syst. (ICCAS)*, Nov. 2022, pp. 1660–1665.

[10] S. Koike, "Performance analysis of signed correlation algorithm with recursive estimation of signed data covariance for fast convergent and robust adaptive filters," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2011, pp. 625–628.

[11] E. Eweda, "Tracking analysis of the sign-sign algorithm for nonstationary adaptive filtering with Gaussian data," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1375–1378, May 1997.

[12] V. J. Mathews, "Performance analysis of adaptive filters equipped with the dual sign algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 85–91, May 1991.

[13] J. Ni, X. Chen, and J. Yang, "Affine projection sign subband adaptive filter," in *Proc. Int. Conf. Image Anal. Signal Process.*, Nov. 2012, pp. 1–4.

[14] Y. Yan, E. E. Kuruoglu, and M. A. Altinkaya, "Adaptive sign algorithm for graph signal processing," *Signal Process.*, vol. 200, Nov. 2022, Art. no. 108662.

[15] N. H. Nguyen, K. Doğançay, and W. Wang, "Adaptive estimation and sparse sampling for graph signals in alpha-stable noise," *Digit. Signal Process.*, vol. 105, Oct. 2020, Art. no. 102782.

[16] N. N. Siqueira, M. R. Petraglia, and D. B. Haddad, "A novel heuristic for the analysis of block sign LMS algorithm," *Electron. Lett.*, vol. 59, no. 7, Apr. 2023.

[17] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[18] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.

[19] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y.-F. Huang, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2928–2941, Aug. 1999.

[20] M. V. S. Lima and P. S. R. Diniz, "Fast learning set theoretic estimation," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2013, pp. 1–5.

[21] R. C. de Lamare and P. Diniz, "Set-membership adaptive algorithms based on time-varying error bounds for CDMA interference suppression," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 644–654, Feb. 2009.

[22] M. V. S. Lima and P. S. R. Diniz, "On the steady-state MSE performance of the set-membership NLMS algorithm," in *Proc. 7th Int. Symp. Wireless Commun. Syst.*, Sep. 2010, pp. 389–393.

[23] S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Process. Lett.*, vol. 8, no. 8, pp. 231–235, Aug. 2001.

[24] M. Z. A. Bhotto and A. Antoniou, "A set-membership nlms algorithm with adaptive error bound," in *Proc. Can. Conf. Electr. Comput. Eng.*, May 2008, pp. 2007–2010.

[25] H. Yazdanpanah and P. S. R. Diniz, "New trinion and quaternion set-membership affine projection algorithms," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 2, pp. 216–220, Feb. 2017.

[26] F. Huang, J. Zhang, and S. Zhang, "Combined-step-size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 5, pp. 493–497, May 2016.

[27] E. Eweda, "Transient performance degradation of the LMS, RLS, sign, signed regressor, and sign-sign algorithms with data correlation," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 46, no. 8, pp. 1055–1062, Jan. 1999.

[28] E. Eweda, "Tracking analysis of the sign algorithm without the Gaussian constraint," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 1, pp. 115–122, Jun. 1998.

[29] E. Eweda, "Almost sure convergence of a decreasing gain sign algorithm for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1669–1671, Oct. 1998.

[30] D. B. Haddad, M. R. Petraglia, and A. Petraglia, "A unified approach for sparsity-aware and maximum correntropy adaptive filters," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 170–174.

[31] N. N. Siqueira, T. P. Ferreira, L. C. Resende, R. M. S. Pimenta, D. B. Haddad, and M. R. Petraglia, "Transient analysis of the set-membership LMS algorithm," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1298–1302, Apr. 2021.

[32] T. R. P. Gonzaga, V. d. S. N. Junior, A. L. F. de Barros, and D. B. Haddad, "Transient analysis of the block least mean squares algorithm," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 608–612, Feb. 2021.

[33] M. D. S. Vieitos, M. P. Tcheou, D. B. Haddad, and M. H. C. Dias, "Improved proportionate constrained normalized least mean square for adaptive beamforming," *Circuits, Syst., Signal Process.*, vol. 42, no. 12, pp. 7651–7665, Aug. 2023.

[34] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.

[35] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.

[36] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, Sep. 2009.

[37] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1428–1439, Jun. 1996.

[38] K. Pärlin, T. Riihonen, V. Le Nir, and M. Adrat, "Estimating and tracking wireless channels under carrier and sampling frequency offsets," *IEEE Trans. Signal Process.*, vol. 71, pp. 1053–1066, 2023.

[39] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Boston, MA, USA: McGraw-Hill, 2002.

[40] T. T. P. Silva, P. Lara, F. Igreja, F. D. V. R. Oliveira, L. Tarrataca, and D. B. Haddad, "An exact expectation model for the LMS tracking abilities," *IEEE Trans. Signal Process.*, vol. 68, pp. 5882–5893, 2020.

[41] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2811–2825, Dec. 1993.

[42] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. IT-4, no. 2, pp. 69–72, Jun. 1958.

[43] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ, USA: Wiley, Oct. 2011.

[44] S. J. M. de Almeida, M. H. Costa, and J. C. M. Bermudez, "A stochastic model for the deficient length pseudo affine projection adaptive algorithm," in *Proc. 17th Eur. Signal Process. Conf.*, Aug. 2009, pp. 1715–1719.

[45] D. Xu, B. Yin, W. Wang, and W. Zhu, "Variable tap-length LMS algorithm based on adaptive parameters for TDL structure adaption," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 809–813, Jul. 2014.

[46] P. Lara, L. D. T. J. Tarrataca, and D. B. Haddad, "Exact expectation analysis of the deficient-length LMS algorithm," *Signal Process.*, vol. 162, pp. 54–64, Sep. 2019.

[47] M. V. S. Lima, P. S. R. Diniz, and H. Yazdanpanah, "Set-membership constrained frequency-domain algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 2, pp. 797–801, Feb. 2021.

[48] *5, Digital Network Echo Cancellers*, document G. 168, ITU-T, 2004.

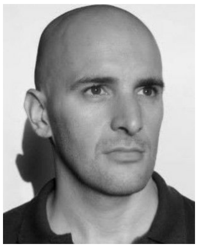
[49] C. G. Lopes, V. H. Nascimento, and L. F. O. Chamon, "Distributed universal adaptive networks," *IEEE Trans. Signal Process.*, vol. 71, pp. 1817–1832, 2023.

[50] X. Huang, Y. Li, Y. Zakharov, Y. Miao, and Z. Huang, "Squared sine adaptive algorithm and its performance analysis," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 3617–3628, 2023.

JOSÉ V. G. DE SOUZA received the B.Sc. degree in control and automation engineering and the M.Sc. degree from the Federal Center for Technological Education, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. He works as a Maintenance Engineer with Petrobras Transporte S.A. His research interests include signal processing, adaptive filtering, adaptive control, and cooperative control.

FELIPE DA R. HENRIQUES received the degree in electrical engineering with emphasis in telecommunications and the master's degree in electronic engineering from the State University of Rio de Janeiro, in 2006 and 2010, respectively, and the Ph.D. degree in electrical engineering from the Federal University of Rio de Janeiro, in 2015. He has experience in electrical engineering, with emphasis on telecommunications systems, signal processing, and computer networks. He is currently the Director of Uned Petrópolis, CEFET/RJ, where he is also an EBTT Professor of technical course in telecommunications integrated into High School and in computer engineering course. In addition, he is with the Graduate Program in Computer Science (PPCIC) and the Graduate Program in Instrumentation and Applied Optics (PPGIO), CEFET/RJ. His research interests include wireless sensor networks, sparse representations and compressive sampling, and adaptive filtering.

NEWTON N. SIQUEIRA was born in Rio de Janeiro, Brazil, in 1979. He received the B.Sc. degree in telecommunication engineering, in 2001, and the M.Sc. and D.Sc. degrees in electrical engineering from the Federal University of Rio de Janeiro, Brazil, in 2006 and 2023, respectively. He is currently with the Federal Center for Technological Education (CEFET/RJ). His research interests include signal processing, machine learning, computer vision, and adaptive filtering algorithms.



LUÍS TARRATACA received the B.Sc., M.Sc., and Ph.D. degrees from Instituto Técnico/Technical University of Lisbon, in 2007, 2008, and 2013, respectively. He is currently a Professor with the Department of Computer Engineering, CEFET-RJ, and a Researcher with the Quantum Computing Group, LNCC, Rio de Janeiro. His research interests include quantum computation, quantum information, machine learning, computer vision, and artificial intelligence algorithms.

FABIO A. A. ANDRADE (Senior Member, IEEE) received the Ph.D. degree in engineering cybernetics from the Norwegian University of Science and Technology. He was an Adjunct Professor with the Federal Center for Technological Education of Rio de Janeiro and a Technical Advisor with TracSense. He was with Brazilian Department of Airspace Control and was an Assistant Professor with Brazilian Naval Academy. He is currently an Associate Professor with the University of South-Eastern Norway and a Research Scientist with the NORCE Norwegian Research Centre. He is the Co-Founder of Umaker. He has published several works on robotics and machine learning. He is the former Chair of the IEEE Norway Chapter of Robotics and Control Systems.

DIEGO B. HADDAD (Member, IEEE) was born in Niterói, Rio de Janeiro, Brazil, in 1983. He received the B.Sc. degree in electrical engineering, in 2005, and the M.Sc. and D.Sc. degrees in electrical engineering from the Federal University of Rio de Janeiro, Brazil, in 2008 and 2013, respectively. He is currently with the Federal Center for Technological Education (CEFET/RJ). His research interests include signal processing, machine learning, computer vision, and adaptive filtering algorithms.